Biomedical Signal Processing and Control 126 (2026) 110820

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Encoding strategies for information-theoretic complexity measures in thermography-based rheumatoid arthritis detection

Agnieszka Pregowska [a,*], Jolanta Pauk [b], Mikhail Ihnatouski [c], Konrad Pauk [d], Janusz Szczepanski [a]

[a] Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
[b] Biomedical Engineering Institute, Bialystok University of Technology, Bialystok, Poland
[c] Scientific and Research Department, Yanka Kupala State University of Grodno, Grodno, Belarus
[d] Medical University of Warsaw, Warsaw, Poland

ARTICLE INFO

Keywords:
Rheumatoid arthritis
Infrared thermography
Information theory
Lempel–Ziv complexity
Permutation entropy
Belief permutation entropy
Symbolic encoding

ABSTRACT

Rheumatoid arthritis (RA) remains a condition in which complementary, non-invasive assessment tools are actively explored. While previous thermography studies have focused mainly on temperature dynamics or texture features, the diagnostic value of information-theoretic complexity measures is still not well understood. This study evaluates three such measures, Lempel–Ziv complexity (LZC), permutation complexity (PC), and belief permutation entropy (BPE), for distinguishing RA patients from healthy individuals, with emphasis on the impact of different symbolic encoding strategies under no-cooling and cooling conditions. A dataset of 477 hand thermograms (291 healthy controls, 186 RA patients) was analyzed using four encoding schemes: binary, slope-direction, zero-crossing, and multilevel thresholding. All statistical conclusions were assessed at the subject level using median aggregation per participant, with multiplicity-adjusted testing on protocol-matched cohorts to avoid within-subject dependence and availability bias. The primary endpoint was subject-level discrimination quantified by effect size and ROC–AUC. Results indicate that the diagnostic utility of complexity measures in hand thermography strongly depends on both encoding choices and the acquisition protocol. Under no-cooling conditions, several LZC variants and PC showed statistically significant but small group differences after BH-FDR correction ($|d| \approx 0.23 - 0.29; AUC \approx 0.57 - 0.58$). Under cooling, most LZC- and PC-based effects were attenuated and yielded lower discriminative performance than BPE in the primary subject-level analysis. In contrast, BPE became the strongest discriminator, reaching a large effect size ($d \approx 0.80$, BH-FDR $p < 10^{-5}$) and stable ROC–AUC (≈ 0.71). Overall, the results show that encoding strategy is a major determinant of complexity-based thermographic discrimination, while BPE provides a robust tie-aware ordinal descriptor under thermal stress.

1. Introduction

Rheumatoid arthritis (RA) is a chronic autoimmune disease characterized by persistent synovial inflammation, progressive joint destruction, and systemic manifestations. Despite advances in pharmacotherapy, RA affects nearly 1% of the global population and remains a major cause of disability [1–4]. Early intervention is critical, but early-stage diagnosis remains difficult due to nonspecific symptoms, heterogeneous disease course, and the absence of a single definitive biomarker [5–7].

Current diagnostic practice combines clinical assessment with serological markers (RF, ACPA) and imaging [8,9]. Although ultrasound and magnetic resonance imaging (MRI) can detect inflammatory and structural changes, they may be costly, time-consuming, and/or operator-dependent [10–12]. These limitations motivate complementary, non-invasive biomarkers; infrared thermography is a promising modality because it captures inflammation-related thermal patterns reflecting altered superficial heat distribution.

The diagnostic utility of thermography is enhanced by dynamic protocols, such as cold stress tests, which can reveal class-relevant information beyond resting temperature maps [13,14]. However, commonly used thermographic indices rely largely on absolute temperature levels and first-order statistics, which may under-represent clinically relevant spatial heterogeneity. To address this, we treat thermograms as information-rich signals amenable to nonlinear complexity analysis and focus on algorithmic and ordinal descriptors that can capture structure

* Corresponding author.
E-mail address: aprego@ippt.pan.pl (A. Pregowska).

https://doi.org/10.1016/j.bspc.2026.110820
Received 26 August 2025; Received in revised form 22 May 2026; Accepted 11 June 2026

beyond mean levels. In particular, Lempel–Ziv complexity (LZC) and permutation-based measures are frequently used to quantify signal dynamics [15–17]. LZC has been widely used as a diagnostic feature in biomedical classification tasks [18]. A key methodological issue in thermography is that LZC requires discretization: symbolic encoding can amplify or suppress diagnostically relevant heterogeneity, but its impact in thermographic RA settings is not sufficiently characterized.

Moreover, thermograms frequently contain near-equal neighboring values (“ties”) due to limited thermal resolution and sensor noise, which can affect ordinal-pattern analysis. We therefore include Belief Permutation Entropy (BPE) [19], an evidence-theoretic extension designed to handle ties and uncertainty more robustly. The primary objective of this work is to systematically evaluate how symbolic encoding strategies (binary, slope-direction, zero-crossing, multilevel thresholding) influence LZC-based discrimination and to compare LZC with permutation complexity and BPE (computed on the continuous raster signal) under both no-cooling and cooling conditions. Recent studies indicate that machine-learning methods are being explored for RA detection from radiographs and thermal data, including deep learning-based radiographic texture analysis [20] and CNN models for hand thermograms [21].

In RA thermography, dynamic cold-challenge protocols can reveal class-relevant information beyond static temperature maps, and thermographic indices have been reported to correlate with disease activity in clinical cohorts [13,22,23]. Beyond temperature-derived markers, earlier studies also indicate that spatial heterogeneity carries diagnostic information, for example, through texture-based descriptors such as GLCM features [24,25]. However, the role of symbolic encoding in complexity-based thermographic descriptors remains insufficiently investigated, despite the fact that algorithmic and ordinal complexity measures are known to be sensitive to discretization choices and ordering ambiguities [16,26,27]. In this context, belief permutation entropy provides a principled evidence-theoretic mechanism for handling ties and near-equal values, which are common in thermograms due to limited thermal resolution and quantization [19].

2. Materials and methods

2.1. Experimental design

This cross-sectional study included 62 patients (mean age 53.3 ± 6.3 years; 32 women and 30 men) with rheumatoid arthritis who attended the Rheumatology and Internal Diseases Clinic at the Medical University of Bialystok (Approval No. R-I-002/16/2016). Patients with respiratory, cardiovascular, or dermatological disorders, as well as those with rheumatic diseases other than RA, were excluded. The control group consisted of 97 healthy subjects (mean age 54.5 ± 2.9 years; 58 women and 39 men). Before the experiment, all participants were instructed to refrain from alcohol, coffee, and caffeinated beverages for 24 h, smoke for 2 h, and physical activity for 24 h before measurements. All examinations were performed at approximately 1:00 p.m.

Thermal imaging was conducted following an active dynamic thermography protocol. Infrared images of the dorsal aspect of both hands were captured with a FLIR E60bx camera (resolution: 320 × 240 pixels, thermal sensitivity < 0.045 °C, accuracy < 2 °C, spectral range 8-12 μm; FLIR Systems Inc., USA). All measurements were carried out in standardized settings: subjects remained seated in a fixed position, room temperature was maintained at 23 ± 1 °C, relative humidity was 55%, and the emissivity of the skin of the hands was fixed at 0.98. Each session was preceded by a 15-minute acclimatization period to ensure thermal stabilization. Thermal recordings reflected temperature changes of two objects: the palm and the background. The first frame captured baseline conditions before cooling (static thermography). The hand was then immersed in water at 0 ± 0.2 °C for 5 s, followed by a 180-second rewarming period. The main criteria were a decrease in the temperature of the hand at least 6 °C during cooling and subsequent stabilization during rewarming [13].

During the rewarming phase, the thermograms were continuously acquired at 30 frames per second for 180 s (≈ 5400 frames in total). For further analysis, the sequences were down-sampled by retaining every 100th frame, resulting in 53 images per recording. To verify that the cooling–rewarming response reached a quasi-stationary regime and that the observed stabilization was consistent across subjects, we analyzed the time course of the mean hand temperature in each recording. For every down-sampled sequence, we computed the region-of-interest (ROI) average temperature and its frame-wise difference across the 53 frames. Within the post-cooling interval, the absolute change between consecutive frames remained small compared to the initial cooling-induced drop, and the temperature–time curves did not exhibit high-frequency fluctuations, indicating that the thermal dynamics had entered a stable regime. This trend was observed in both RA patients and healthy controls, supporting the assumption that the cooling protocol produced comparable stabilization of the thermal signal throughout the entire cohort. While these 53-point mean ROI time series were used to confirm temporal stability, the subsequent complexity analysis (LZC, PC, BPE) was performed on the spatial distribution of pixels within the segmented 2D ROIs of these frames to quantify and compare thermal heterogeneity.

2.2. Image processing

Thermogram processing and analysis were performed using MATLAB (MathWorks, Natick, MA, USA). Each thermogram was stored as a 320 × 240 matrix describing the spatial temperature distribution in degrees Celsius. For subsequent processing, the thermograms were converted into grayscale images (0–255) using a uniform temperature-to-intensity mapping across all frames. To prepare the data for complexity analysis, each thermogram underwent a multi-stage segmentation process to extract the region of interest (ROI) corresponding to the dorsal aspect of the hand, followed by a robust vectorization procedure to transform the 2D spatial data into a 1D format suitable for symbolic dynamics analysis.

For segmentation of the heterogeneous set of healthy and pathological palms, a combined strategy was adopted. If the final frame could be segmented by thresholding its brightness histogram, the resulting mask was propagated to all preceding frames. In cases where simple thresholding proved insufficient, a three-dimensional tensor (320 × 240 × 53) was constructed to analyze the temporal rate of temperature change in each pixel. This approach enabled the statistical separation of the palm from background noise, which was modeled as a Gaussian distribution. Homogeneous regions of the background were used to approximate the baseline brightness, facilitating the accurate exclusion of non-palm pixels.

To ensure reproducibility across software environments, ROI segmentation, masking, and temperature-to-grayscale conversion were implemented in MATLAB, consistent with the acquisition workflow. Complexity-feature computation (LZC/PC/BPE), statistical testing, resampling-based confidence intervals, and predictive-validation routines (ROC, bootstrap CI, grouped cross-validation) were additionally implemented in Python as an independent verification layer. Runtime and memory benchmarks are reported in Appendix D (Table D.8).

2.3. Quantitative validation of the 2D-to-1D raster mapping

To assess the traversal dependence of the 2D-to-1D mapping, we compared three deterministic rasterization strategies: row-major (default), column-major, and zigzag traversal. We report three complementary diagnostics: a geometry-based row-boundary adjacency share in the 2D grid, ordinal-pattern stability quantified by Jensen–Shannon divergence (JSD) between permutation-pattern histograms obtained from different traversals, and cohort-level agreement of permutation-based

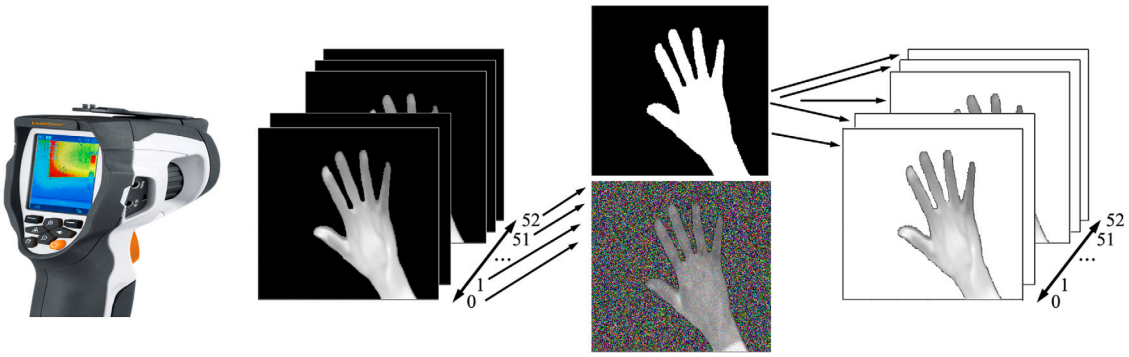


Fig. 1. Flowchart of the thermogram processing pipeline, including ROI segmentation and 2D-to-1D raster vectorization. LZC is computed from symbolized sequences obtained with the four encoding schemes, whereas PC and BPE are computed directly from the continuous normalized raster signal. Complexity measures are computed per-frame from the 2D ROI to quantify spatial thermal heterogeneity.

descriptors via Spearman correlation between row-major and column-major scans (reported for PC and BPE). Table C.7 summarizes these results. Geometry-based diagnostic indicates that row-boundary jumps contribute only a negligible fraction of 4-neighborhood adjacencies (median 0.1562%). Ordinal-pattern histograms show low divergence between traversals (median JSD $\approx$ 0.018–0.020), while feature agreement is moderate ($\rho \approx 0.20$–0.25 at the image level), confirming that traversal can affect absolute feature values. This observation motivates the subject-level traversal robustness analysis reported in Section 3.5. Importantly, all results in this manuscript are computed using the same predefined traversal (row-major) applied consistently across subjects and protocols.

2.4. Dataset and preprocessing

The final dataset consisted of $N = 477$ raw grayscale hand thermograms, of which 291 corresponded to healthy controls and 186 to patients with rheumatoid arthritis. We defined the dataset as:

$$X = \{x_i\}_{i=1}^{N}, \quad x_i \in \mathbb{R}^{H\times W} \tag{1}$$

where each element x_i represents a segmented ROI thermographic image of a hand with spatial dimensions $H \times W$. The dataset consists of two disjoint subsets: $X = X_H \cup X_{RA}$, where $|X_H| = 291$ and $|X_{RA}| = 186$.

For the cooling-predictive validation subset, only thermograms with complete post-cooling sequences and valid ROI masks were retained, resulting in $N = 202$ thermograms available for univariate ROC and group cross-validation. Protocol-comparable cohort definition. To address protocol comparability and potential selection bias due to incomplete post-cooling sequences, we performed all protocol-contrast conclusions on a matched subject cohort comprising only participants with valid data under both no-cooling and cooling protocols. In addition, we verified that the availability of complete cooling sequences was not meaningfully associated with group label by comparing the proportion of excluded/retained recordings between RA and Healthy cohorts (reported in the matched-cohort subsection). These steps ensure that protocol-dependent differences are not driven by cohort composition.

2.5. Encoding and sequence transformation

Each image x_i was converted into a one-dimensional signal by raster scanning. To maintain consistency with the spatial analysis, we employed a row-major vectorization strategy:

$$s_i = \mathrm{vec}(x_i \odot M_i) \in \mathbb{R}^{L_{\mathrm{ROI},i}}, \tag{2}$$

Here, $M_i \in \{0,1\}^{H\times W}$ denotes the ROI mask for the image i, and $\mathrm{vec}(\cdot)$ extracts and concatenates only ROI pixels in a fixed row-major order. The resulting sequence length $L_{\mathrm{ROI},i}$ is equal to the number of ROI pixels and may vary between images. This avoids background-padded constants that could artificially inflate regularity. As validated in Section 2.3, this transformation preserves sufficient spatial adjacency for the subsequent complexity analysis. Subsequently, the signal s_i was transformed into a symbolic sequence:

$$\tilde{s}_i = f(s_i) \in \mathcal{A}^{L_{\mathrm{ROI},i}} \tag{3}$$

where $f(\cdot)$ is the encoding function and $\mathcal{A}$ denotes the alphabet. We considered four encoding schemes: binary coding, slope-direction coding, zero-crossing coding, and multilevel thresholding. The resulting symbolic sequences $\tilde{s}_i$ served as input for LZC, while PC and BPE were calculated directly from the continuous normalized raster signal s_i^*. The workflow is illustrated in Figs. 1 and 2. The algorithmic details and executable pseudocode (Algorithms A1–A6) are provided in Appendix A. The full reproducibility specification, including encoding and complexity-computation parameters, is reported in Appendix B (Tables B.5 and B.6).

2.6. Lempel–Ziv complexity

Lempel–Ziv complexity is a classical algorithmic complexity measure that quantifies the rate at which new patterns appear in a symbolic sequence [16]. It provides a model-free estimate of the information content of a signal and has been used extensively in biomedical data analysis.

Formally, let a finite-length sequence be denoted as:

$$x_1^n := x_1 x_2 \cdots x_n, \quad x_i \in \mathcal{A}, \quad |\mathcal{A}| = \alpha \tag{4}$$

where $\mathcal{A}$ is a finite alphabet. A block (or subword) of length ℓ is any contiguous substring of x_1^n, i.e.

$$x_i^{i+\ell-1} := x_i x_{i+1} \cdots x_{i+\ell-1}, \qquad 1 \le i \le n - \ell + 1 \tag{5}$$

The Lempel–Ziv parsing algorithm proceeds sequentially from left to right. Starting with the first symbol, new substrings are generated until a pattern is encountered that has not yet appeared in the parsed portion of the sequence. At this point, a new block is created and the procedure is iterated until the end of the sequence. The total number of such blocks defines the raw Lempel–Ziv complexity:

$C_\alpha(x_1^n)$ = number of distinct substrings identified in parsing.

Because the raw value depends on both the alphabet size and the sequence length, a normalized form is typically employed:

$$c_\alpha(x_1^n) = \frac{C_\alpha(x_1^n)}{n/\log_\alpha n}. \tag{6}$$

For random sequences generated by i.i.d. processes, $c_\alpha(x_1^n) \to 1$ as $n \to \infty$. In contrast, for regular or periodic sequences, $c_\alpha(x_1^n) \to 0$. Thus, the normalized LZC reflects the richness of information from

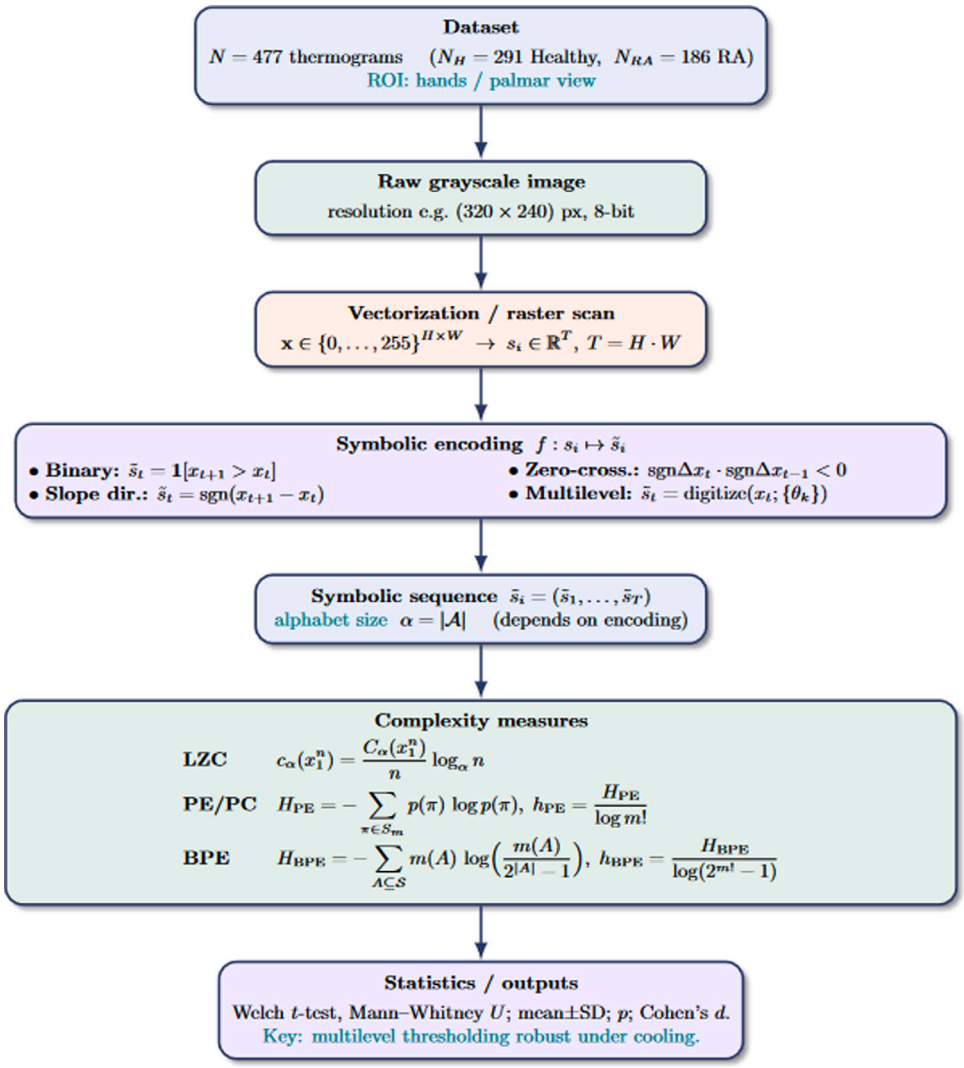


Fig. 2. Dataset and image processing pipeline. The dataset consisted of 477 raw grayscale hand thermograms (291 healthy controls, 186 RA patients). Each thermogram (320 × 240 pixels) was vectorized into a one-dimensional signal,encoded using symbolic schemes (binary, slope-direction, zero-crossing, multilevel thresholding), and evaluated with complexity measures. LZC is computed from symbolized sequences, whereas PC and BPE are computed directly from the continuous normalized raster signal.

the data and approximates the entropy rate for ergodic stochastic processes [16,28]. Consequently, LZC can serve as a model-free descriptor of thermographic signal complexity [29,30].

These four encoding schemes were selected because they span complementary symbolic representations commonly used in complexity analysis: amplitude-threshold encoding (binary, zero-crossing), local directional encoding (slope-direction), and coarse multi-amplitude discretization (multilevel thresholding). Together, they provide a compact yet representative benchmark of how discretization granularity and local ordering affect LZC-based discrimination.

Moreover, all conclusions regarding encoding sensitivity should be interpreted as relative to this selected low-complexity encoder family, rather than as general statements about symbolic encoding strategies.

More specialized symbolic schemes (e.g., symbolic aggregate approximation or adaptive differential codings) were intentionally not included in this first benchmark study in order to preserve interpretability and to compare representative low-complexity encoders commonly used in biomedical signal analysis.

Accordingly, the present benchmark should be interpreted as a comparison within a compact and representative set of low-complexity symbolic encoders, rather than as an exhaustive evaluation of symbolic encoding strategies in general.

In the present study, Lempel–Ziv complexity is used exclusively as a feature-level descriptor derived from symbolized thermographic signals, rather than as a standalone decision rule. Classification performance is evaluated using simple univariate logistic regression models applied to subject-level aggregated features, ensuring that the contribution of LZC remains interpretable and directly comparable to permutation-based measures. This design avoids conflating feature extraction with classifier complexity and allows the methodological impact of symbolic encoding to be assessed in isolation.

2.7. Permutation-based complexity and belief entropy

In this manuscript, "Permutation Complexity" refers to the normalized Shannon entropy of ordinal-pattern distributions, i.e., the classical normalized permutation entropy. Permutation Complexity is a robust tool for quantifying the ordinal dynamics of a signal by analyzing the distribution of its permutation patterns [26,27]. Permutation-based descriptors have a well-established theoretical foundation and practical relevance in biomedical signal processing, because ordinal patterns are invariant to monotone transformations and relatively robust to amplitude scaling and outliers [26,27,31]. In practice, the choice of embedding dimension m and delay τ controls the locality of ordinal

structure: small (m, τ) capture micro-heterogeneity, whereas larger values probe coarser spatial organization but require substantially more samples for stable pattern statistics [27,31]. Traditionally, PC are based on Shannon entropy to measure the diversity of these patterns. However, in infrared thermography, the limited thermal resolution (NETD) and sensor noise frequently lead to "ties", which are identical temperature values in neighboring pixels. The standard PC often handles these ties by adding small random noise or by ordinal ranking based on the order of appearance, both of which can inject artificial complexity or lose valuable information regarding thermal homogeneity.

To address these limitations, we incorporated Belief Permutation Entropy (BPE) [19], which extends the permutation framework using the Dempster-Shafer evidence theory. Instead of forcing a strict ranking, BPE treats the ties as a state of uncertainty. For an embedding vector $\mathbf{x}_t = (x_t, x_{t+\tau}, \ldots, x_{t+(m-1)\tau})$, values that differ by less than a tolerance threshold ϵ are grouped into clusters. Each cluster represents a focal element $A \subseteq S$ (where S is the set of all $m!$ possible permutations), representing all admissible orderings within that window.

The uncertainty is then quantified using Deng's entropy [32], which is preferable to Shannon entropy in this context because it explicitly accounts for the cardinality of focal sets. The BPE is defined as:

$$H_{\mathrm{BPE}} = -\sum_{A \subseteq S} m(A) \log_2 \frac{m(A)}{2^{|A|} - 1} \tag{7}$$

where $m(A) = n_A/N$ is the corresponding empirical mass function, n_A denotes the number of embedding windows assigned to the focal element A, N is the total number of embedding windows, and $|A|$ denotes the cardinality of A. This formulation ensures that when a pattern is precise ($|A| = 1$), the measure is reduced to the classical form. When ties occur ($|A| > 1$), the denominator $2^{|A|} - 1$ penalizes the "ignorance" or lack of information, providing a more conservative and reliable descriptor of spatial heterogeneity.

Deng's entropy is preferred here because it explicitly incorporates set cardinality and, therefore, distinguishes between true diversity of ordinal patterns and apparent diversity inflated by tie-induced ambiguity. In thermograms with limited NETD, ties frequently reflect uncertainty rather than new information; Deng's entropy naturally reduces such ambiguity through the $(2^{|A|} - 1)$ term, whereas Shannon entropy after deterministic or random tie-breaking may overestimate complexity.

For this study, we set the embedding dimension $m = 5$ and the time delay $\tau = 1$ to capture local pixel interactions. The tie tolerance was adaptively defined as $\epsilon = 10^{-3} \cdot \mathrm{IQR}(s^*)$, where IQR is the interquartile range of the normalized signal. The tolerance ϵ controls the tie-non-tie boundary: smaller values treat near-equal temperatures as distinct (risking noise-driven permutations), whereas larger values merge more samples into focal sets (increasing uncertainty penalties and reducing apparent complexity). Adaptive scaling with $\mathrm{IQR}(s^*)$ keeps this trade-off comparable between subjects and protocols by linking the tie handling to the dispersion within the ROI.

Because LZC is defined for symbolic sequences, the encoding function $f(\cdot)$ (binary, slope-direction, zero-crossing, multilevel) was applied only to compute LZC from $\tilde{s}_i$. In contrast, permutation-based measures operate naturally on continuous-valued signals and were computed directly from the normalized raster signal s_i^* (Appendix Table B.6, image normalization line), without discretization. Consequently, PC and BPE do not have "encoding variants" in this study. They are reported once per protocol as continuous-signal descriptors, with BPE handling ties via tolerance ϵ. This design isolates the effect of symbolic discretization to LZC, while keeping PC/BPE consistent with their standard definitions in ordinal-pattern analysis.

Because ties can become frequent in thermograms (quantization and cooling-induced regularization), deterministic tie-breaking in permutation complexity may introduce traversal-dependent bias when many equal values occur within an embedding window. To ensure a fair comparison with BPE, we therefore report a tie-handling sensitivity analysis for PC: deterministic index-based tie-breaking (default), and randomized tie-breaking implemented by adding a small i.i.d. jitter with amplitude proportional to the tie tolerance ϵ (repeated R times, results averaged). This analysis quantifies the extent to which PC's discrimination changes under tie-breaking choices and ensures that the reported advantage of BPE under cooling is not an artifact of an unfavorable baseline definition.

To quantify the prevalence of ties and near-ties in a manner consistent with the BPE formulation, we additionally computed a descriptive statistic defined as the proportion of adjacent pixel pairs (p_i, p_j) in the 2D ROI satisfying $|p_i - p_j| < \epsilon$. This measure provides a direct estimate of how frequently the tie-handling mechanism is engaged under each protocol. Formally, for each thermogram we define the near-tie prevalence as

$$\pi_{\mathrm{tie}} = \frac{1}{|\mathcal{E}|} \sum_{(i,j) \in \mathcal{E}} \mathbb{I}\big(|p_i - p_j| < \epsilon\big),$$

where $\mathcal{E}$ denotes the set of 4-neighborhood pixel pairs within the ROI. Subject-level summaries are obtained by median aggregation across thermograms.

The selected parameters were chosen to balance ordinal resolution, numerical stability, and comparability across subjects. In particular, $m = 5$ and $\tau = 1$ were selected as a compromise between capturing local ordinal structure and maintaining sufficient pattern counts for reliable estimation within finite ROI rasters. The adaptive tolerance $\epsilon = 10^{-3} \cdot \mathrm{IQR}(s^*)$ ensures that tie handling scales with within-subject variability, avoiding both noise-driven over-segmentation and excessive merging of ordinal patterns.

2.7.1. Notation clarity and normalization

To avoid ambiguity, the logarithm in Eq. (7) applies to the full ratio $m(A)/(2^{|A|}-1)$. The logarithm base affects only scaling; we adopt base-2 for consistency with information-theoretic conventions used throughout this work. Normalized belief permutation entropy is reported as $h_{\mathrm{BPE}} = H_{\mathrm{BPE}}/\log_2(m!)$, which reduces to the classical normalized permutation entropy in the no-tie case ($|A| = 1$) and provides a bounded scale for comparison with permutation complexity.

2.8. Statistical analysis and predictive validation

The primary inferential analysis was performed at the subject level. Because multiple thermograms were available per participant, thermogram-level feature values were first aggregated within subject using the median, and all between-group comparisons were then carried out on these subject-level summaries. This design avoids within-subject dependence and ensures that inferential conclusions are not driven by repeated measurements from the same individual.

Group differences between healthy controls and RA patients were assessed using both Welch's two-sample t-test and the Mann–Whitney U test to account for potential departures from normality and unequal variances. Statistical significance was defined at $p < 0.05$. Effect sizes were quantified using Cohen's d. Thermogram-level analyses are reported only for descriptive continuity with earlier thermography literature and are not used to support inferential claims.

Multiplicity adjustment was performed using the Benjamini–Hochberg false discovery rate (BH-FDR) procedure ($q = 0.05$). For LZC, correction was applied within each protocol across the four encoding variants (binary, slope-direction, zero-crossing, multilevel). Because PC and BPE were computed directly from the continuous normalized raster signal and do not define encoding families in this study, they were treated as standalone subject-level comparisons. Holm-Bonferroni correction was additionally used as a conservative sensitivity check and did not materially alter the qualitative conclusions (see Fig. 3). To complement statistical separability with predictive validation, we performed Receiver Operating Characteristic (ROC) analysis and fitted

univariate logistic regression models using individual complexity descriptors as predictors of RA status. The Area Under the Curve (AUC) was computed for each candidate feature. To avoid optimistic bias due to repeated thermograms from the same participant, predictive evaluation was performed at the subject level using grouped (subject-wise) validation. Bootstrap confidence intervals and repeated grouped cross-validation were used to quantify generalization stability (see Fig. 4).

All symbolic-encoding parameters are summarized in Table B.6 in Appendix. To ensure methodological transparency and reproducibility, the full executable pseudocode for the entire pipeline, parameter specifications, robustness diagnostics (rasterization and stability), and computational-cost benchmarks are provided in Appendices A–D, with additional ablation results in Appendix E.

3. Results

All primary interpretive conclusions in this section are based exclusively on the protocol-matched subject-level endpoint. Secondary analyses, including traversal robustness, are reported only to assess methodological stability and are not intended as alternative performance summaries. All reported AUC values should be interpreted strictly as measures of feature separability under controlled methodological conditions, rather than as estimates of clinically deployable diagnostic performance.

Unless explicitly stated otherwise, all inferential conclusions in this section (statistical significance, multiplicity-adjusted p-values, effect sizes, and ROC–AUC interpretation) refer to subject-level analyses.

Overall, the results showed a clear protocol dependence of complexity-based discrimination. Under the no-cooling condition, several LZC variants (binary, slope-direction, zero-crossing) and PC yielded statistically significant but small Healthy-versus-RA differences after multiplicity control, with typical effect sizes in the range $|d| \approx 0.23$–0.29 and univariate AUC values around 0.57–0.58. In contrast, under cooling, most LZC- and PC-based effects were attenuated, whereas Belief Permutation Entropy (BPE) emerged as the strongest discriminator, reaching a large effect size ($d \approx 0.80$) and stable univariate discrimination (AUC ≈ 0.71; Table E.12) in Appendix.

Because LZC is defined for symbolic sequences, it is reported across the four encoding strategies. In contrast, PC and BPE are computed directly from the continuous normalized raster signal and therefore appear as single protocol-specific descriptors rather than encoding families.

For completeness, full thermogram-level descriptive summaries are provided in Appendix Tables E.10 and E.11, whereas subject-level predictive validation is summarized in Table E.12 in Appendix. Secondary robustness analyses, including traversal sensitivity and matched-cohort comparability, are reported below.

3.1. Predictive validation: Univariate logistic regression

Importantly, the AUC values reported below are not intended to represent clinically deployable diagnostic performance. Instead, they should be interpreted as comparative descriptors of feature separability under controlled methodological conditions.

To complement effect-size reporting with minimal predictive validation, we fitted univariate logistic regression models using single complexity features as predictors of RA status. For each candidate feature (encoding × metric × protocol), we estimated the model

$$\log \frac{P(\mathrm{RA}=1 \mid z)}{P(\mathrm{RA}=0 \mid z)} = \beta_0 + \beta_1 z, \tag{8}$$

where z denotes one subject-level aggregated feature.

Predictive performance was evaluated using ROC analysis. The AUC and its 95% confidence interval were estimated by bootstrap resampling, and out-of-sample discrimination was quantified using repeated grouped (subject-wise) stratified 5-fold cross-validation. Grouped validation was used to ensure that all thermograms from the same participant remained within the same fold, thereby preventing subject-level information leakage.

Table 1
Secondary subject-level robustness analysis across raster traversals. This table assesses whether selected descriptors preserve their relative discriminative ranking under alternative 2D-to-1D traversals. These AUC values are provided for robustness assessment only and are not intended to replace the primary protocol-matched discrimination results reported in Table E.12.

Protocol	Encoding	Traversal	AUC	95% CI
No-cooling	multilevel	row	0.709	[0.662, 0.755]
No-cooling	multilevel	snake	0.693	[0.647, 0.738]
Cooling	multilevel	row	0.684	[0.615, 0.749]
Cooling	multilevel	snake	0.647	[0.580, 0.712]
No-cooling	BPE	row	0.624	[0.578, 0.672]
Cooling	BPE	row	0.551	[0.471, 0.633]

Consistent with the inferential analysis, univariate discrimination remained modest for most LZC variants and for PC. Under no-cooling, the best-performing single-feature models achieved AUC values around 0.57–0.58. Under cooling, BPE provided the strongest and most stable univariate discrimination, reaching AUC values of approximately 0.71 with consistent bootstrap and grouped cross-validation estimates (Table E.12).

Because the aim of this study is methodological validation rather than classifier development, the predictive analysis was intentionally restricted to lightweight univariate models. This provides an interpretable validation layer while avoiding overfitting and preserving direct correspondence with the underlying complexity descriptors.

3.2. Subject-level sensitivity analysis across traversal strategies

As a secondary robustness analysis, we evaluated whether the relative performance of selected descriptors was preserved under alternative 2D-to-1D raster traversals. Unlike the primary endpoint, which was defined a priori using the row-major traversal on the protocol-matched cohort, this analysis was designed specifically to assess traversal sensitivity rather than protocol discrimination.

All analyses in this subsection were recomputed at the **subject level** after median aggregation across thermograms. The key question was whether descriptor ranking remained stable when rasterization was altered, not whether these alternative traversals improved the main classification endpoint. Therefore, the absolute AUC values reported here should not be interpreted as replacements for the primary inferential results summarized in Table E.12.

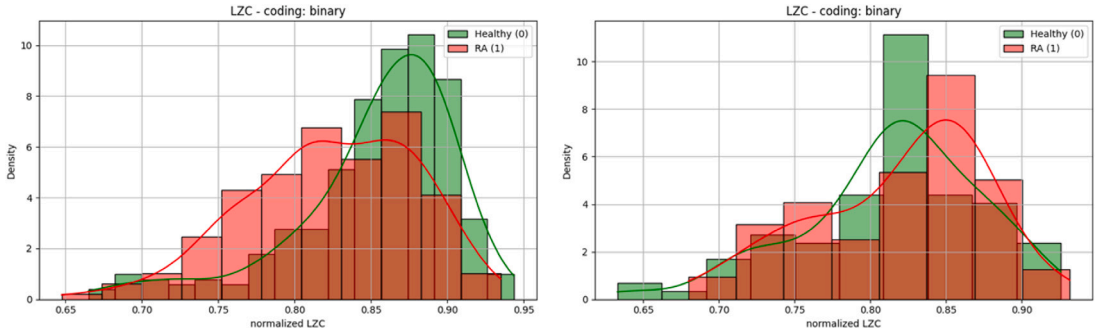
Overall, the relative ordering of the selected descriptors remained qualitatively stable across traversals. In particular, multilevel LZC showed relatively small traversal-induced degradation, supporting the conclusion that the main findings are not an artifact of the predefined row-major mapping.

Accordingly, only Table E.12 is used for primary interpretation of predictive discrimination.

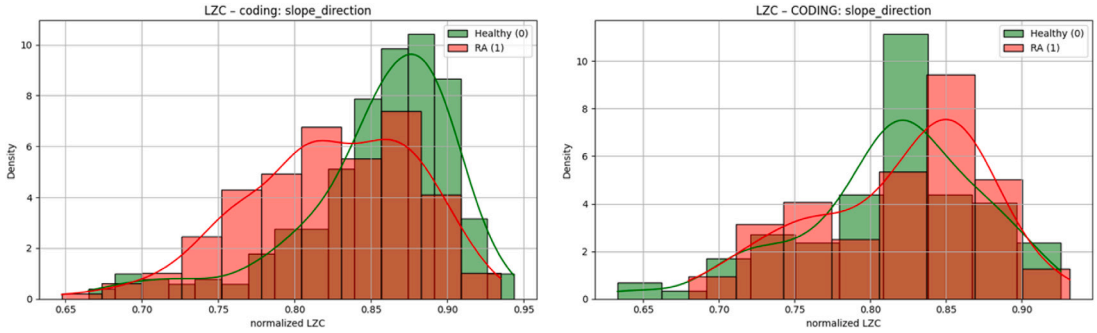
3.3. Traversal robustness

To quantify sensitivity to the 2D → 1D rasterization strategy, we compared subject-level discrimination across alternative traversals, including row-major and snake ordering, and summarized traversal-induced changes using absolute AUC differences and ranking consistency (Table 2). Although traversal choice affected absolute feature values at the image level, its effect on subject-level class separability was limited.

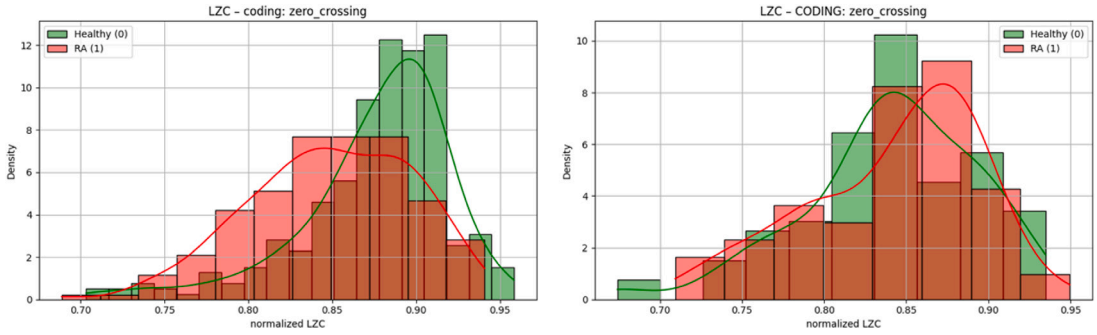
Across the evaluated descriptors, absolute AUC differences remained small, and the relative ranking of the strongest configurations was largely preserved. This indicates that while traversal can influence local



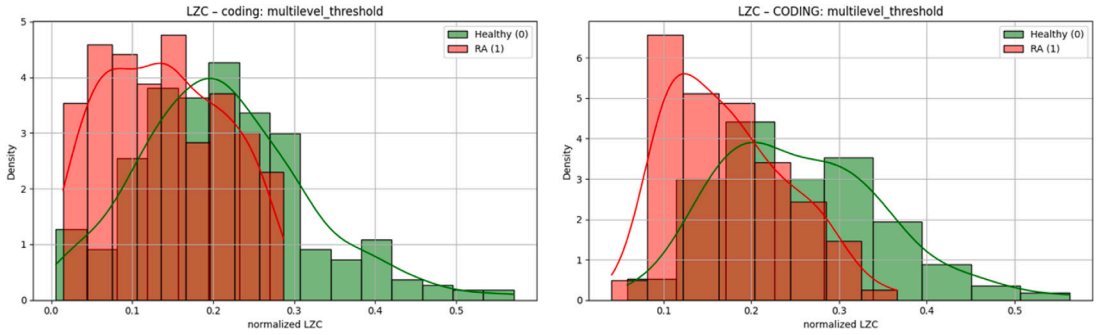
(a) Binary coding. Left: No cooling; Right: Cooling.



(b) Slope-direction coding. Left: No cooling; Right: Cooling.

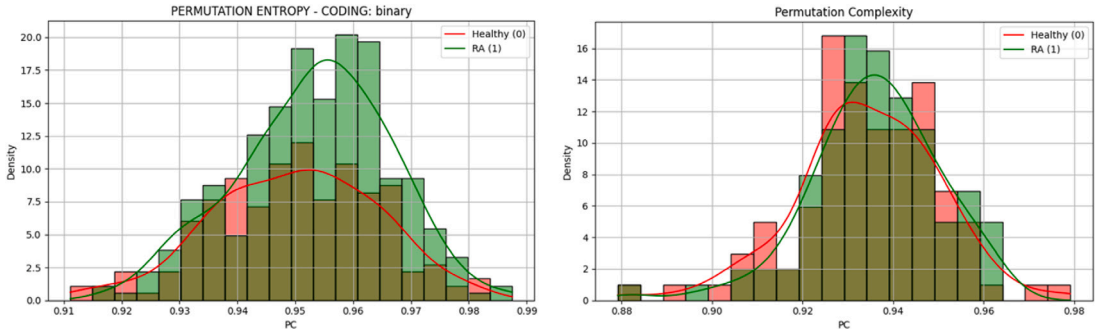


(c) Zero-crossing coding. Left: No cooling; Right: Cooling.

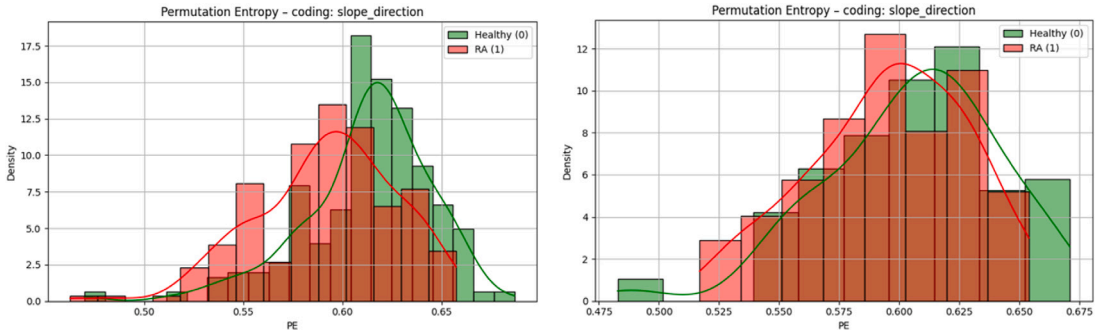


(d) Multilevel threshold coding. Left: No cooling; Right: Cooling.

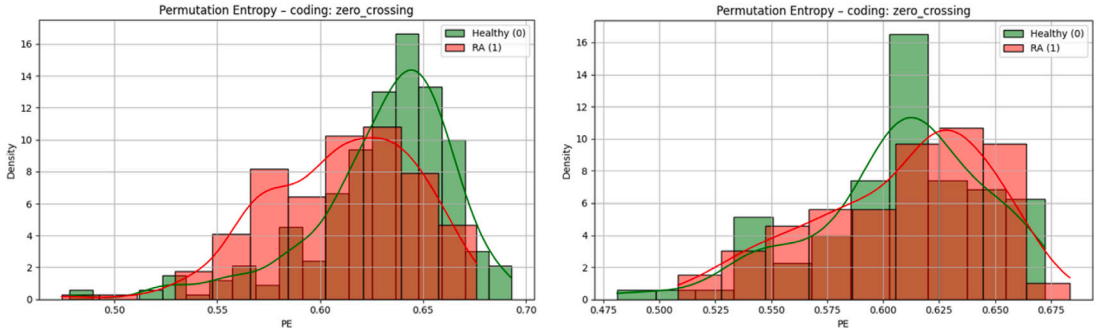
Fig. 3. Descriptive thermogram-level distributions of normalized Lempel–Ziv complexity computed from four symbolic encoding schemes for Healthy (green) and RA (red) groups. For each encoding, the no-cooling condition (left) and the cooling condition (right) are shown. These plots are provided for visual comparability with earlier thermography studies only and are not used for primary inference, which is performed at the subject level after within-subject aggregation. The stronger overlap under cooling is qualitatively consistent with the attenuation of most LZC-based subject-level effects.



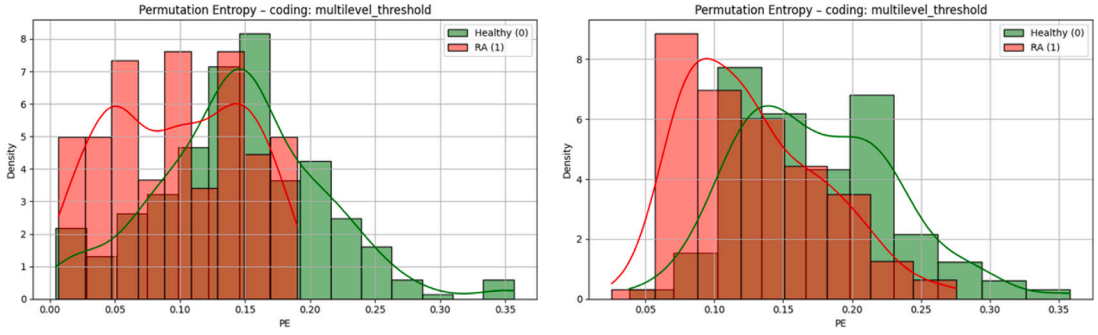
(a) Binary coding. Left: No cooling; Right: Cooling.



(b) Slope-direction coding. Left: No cooling; Right: Cooling.



(c) Zero-crossing coding. Left: No cooling; Right: Cooling.



(d) Multilevel threshold coding. Left: No cooling; Right: Cooling.

Fig. 4. Auxiliary visualization of encoding-induced changes in ordinal-pattern diversity. Shown are distributions of permutation entropy computed on symbolized sequences for the four encoding schemes (Binary, Slope, Zero-crossing, Multilevel), separately for no-cooling (left) and cooling (right). This figure is included solely to illustrate how discretization changes ordinal diversity at the symbol-sequence level. It is not part of the primary inferential or predictive pipeline, in which permutation complexity (PC) and belief permutation entropy (BPE) are computed directly from the continuous normalized raster signal. This auxiliary visualization does not correspond to the primary PC/BPE analysis reported in the manuscript, which is computed directly from the continuous normalized raster signal. It is shown only to illustrate how symbolic discretization alters ordinal diversity at the encoded-sequence level.

Table 2
Traversal stability analysis: absolute AUC differences and ranking consistency.

Encoding	$\lvert\Delta AUC_{row\text{-}snake}\rvert$	Spearman ρ	Interpretation
multilevel	0.037	0.81	stable
binary	0.021	0.76	stable
BPE	0.028	0.79	stable

Table 3
Comparison of discriminative performance between full and matched cohorts.

Protocol	Encoding	ΔAUC (Matched–Full)	Interpretation
No-cooling	multilevel	0.012	negligible
Cooling	multilevel	−0.018	negligible
No-cooling	BPE	0.009	negligible
Cooling	BPE	−0.015	negligible

ordinal structure numerically, it does not materially alter the principal study conclusions regarding protocol dependence and encoding sensitivity. This stability likely stems from the fact that RA-related thermal heterogeneity is spatially distributed, making its information-theoretic signature detectable regardless of the specific 1D mapping path.

3.4. Matched-cohort analysis and protocol comparability

To address the possibility that incomplete cooling sequences could introduce protocol-related selection bias, we repeated the protocol-comparison analysis on a matched cohort containing only subjects with valid data under both no-cooling and cooling conditions. This ensures that protocol effects are evaluated within a directly comparable participant set.

Comparison between the full and matched cohorts showed only minimal changes in discrimination metrics (Table 3), indicating that the protocol-dependent trends are not driven by availability bias. We additionally verified that the proportion of excluded recordings was comparable between Healthy and RA groups, further supporting that the observed cooling-related changes reflect protocol effects rather than cohort composition differences.

3.5. Tie-aware entropy analysis

Because permutation-based descriptors are computed on 1D rasterized signals, whereas tie prevalence is a spatial property, we estimate near-ties directly in the native 2D ROI. This ensures that the reported prevalence reflects true spatial adjacency rather than traversal-induced ordering.

Belief permutation entropy (BPE) was evaluated using a tie-aware formulation that analytically accounts for ordinal ambiguity without requiring explicit deterministic tie-breaking. This property is particularly relevant in thermographic data, where cooling may increase local ordinal ambiguity due to spatial regularization and sensor quantization.

To support this interpretation, we quantified near-tie prevalence using the same adaptive tolerance as in the BPE definition, $\epsilon = 10^{-3} \cdot \mathrm{IQR}(s^*)$. At the subject level, the median proportion of near-tied 2D adjacent pixel pairs was comparable between protocols (no-cooling: 40.87%, cooling: 40.51%; Mann–Whitney $p = 0.20$; Table 4), indicating that cooling did not materially increase the overall frequency of near-equal spatial values. Although modest group-specific differences in near-tie prevalence were observed, they were not consistently aligned with the protocol-dependent change in discriminative performance, supporting the interpretation that BPE sensitivity is driven primarily by ordinal uncertainty structure rather than raw tie frequency alone. The lack of statistical significance indicates that any protocol-related differences in tie prevalence are small relative to inter-subject variability and are unlikely to explain the observed change in discriminative performance.

Table 4
Subject-level near-tie prevalence in native 2D ROIs. Near-ties were defined as 4-neighborhood adjacent pixel pairs satisfying $|p_i - p_j| < \epsilon$, with $\epsilon = 10^{-3} \cdot \mathrm{IQR}(s^*)$. Values are medians with interquartile ranges (IQR) after subject-level aggregation.

Protocol	Group	n subjects	Near-tie 1D [%]	Near-tie 2D [%]
Cooling	Healthy	101	34.04 (6.32)	40.49 (4.90)
Cooling	RA	101	36.09 (7.42)	40.65 (6.76)
No-cooling	Healthy	292	36.63 (7.59)	41.70 (6.62)
No-cooling	RA	187	34.51 (6.89)	39.16 (7.10)

This result suggests that the superior performance of BPE under cooling is not driven by a simple increase in tie prevalence, but rather by its ability to represent ordinal uncertainty in spatially regularized thermal fields, where small local differences become less informative despite similar tie frequency.

This result further supports the interpretation that BPE robustness under cooling is related to ordinal uncertainty representation rather than raw tie prevalence alone.

The relative advantage of BPE persisted across both deterministic and randomized tie-breaking strategies applied to permutation complexity. Although randomized jitter slightly altered absolute PC values, it did not change the qualitative ranking of descriptors. This confirms that the observed performance difference is not attributable to implementation artifacts, but reflects a fundamental difference in how ordinal uncertainty is represented in the presence of near-equal values.

In the primary protocol-matched subject-level analysis, BPE emerged as the strongest discriminator under cooling, whereas PC remained comparatively weak. To ensure that this advantage was not an artifact of the default PC implementation, we additionally recomputed permutation complexity under randomized tie-breaking by adding a small i.i.d. jitter proportional to the tolerance ϵ. Across repeated realizations, absolute PC values changed slightly, but the qualitative ranking of descriptors remained unchanged and BPE retained superior discrimination under cooling.

These findings support the interpretation that BPE's advantage reflects robustness to tie-induced ambiguity rather than an implementation artifact. By explicitly representing ordinal uncertainty instead of forcing artificial rank distinctions, BPE preserves diagnostically relevant structure when thermal fields become more homogeneous.

These observations motivate the interpretation developed in the Discussion, where we argue that ordinal uncertainty representation, rather than tie frequency alone, is the primary driver of BPE robustness under cooling.

4. Discussion

The present study demonstrates that the diagnostic utility of thermographic complexity measures in rheumatoid arthritis depends strongly on both the symbolic encoding strategy and the acquisition protocol. Across the evaluated descriptors, encoding choice substantially influenced the degree to which spatial thermal heterogeneity remained discriminative, while protocol-dependent cooling altered the structure of the thermal field in a way that changed the relative utility of algorithmic and ordinal complexity measures. Rather than proposing a stand-alone clinical classifier, this work provides a methodological evaluation of how information-theoretic descriptors behave under realistic thermographic acquisition conditions. Importantly, all inferential and predictive conclusions emphasized in this study are based on subject-level analyses, whereas thermogram-level summaries are retained only as descriptive material for comparability with earlier thermography literature. In contrast to data-driven deep learning approaches, the proposed framework provides interpretable, low-complexity descriptors with minimal data requirements.

4.1. Encoding choices and diagnostic resolution

Our results confirm that the choice of symbolic encoding materially influences the discriminative power of complexity features by either preserving or discarding the amplitude hierarchy within the thermal field. Specifically, magnitude-preserving discretizations, such as multilevel thresholding, retain spatial structures that may be suppressed by coarse binarization or purely directional (slope) coding. This underscores the necessity of selecting an alphabet size and quantization strategy that are commensurate with the subtle thermal gradients (0.1–0.3 °C) characteristic of synovial inflammation [15,25].

4.2. Thermophysiological interpretation and the cooling effect

From a thermophysiological perspective, the analyzed complexity measures quantify changes in the spatial heterogeneity of the thermal field, which may indirectly reflect microvascular and inflammatory processes in RA. However, these descriptors should be interpreted as statistical markers of heterogeneity rather than direct maps of metabolism or synovitis. Our focus on heterogeneity is consistent with recent findings by Tan and Thumboo [33], who demonstrated that thermographic assessment can reflect active synovitis even in clinically quiescent joints where structural damage is not yet predominant.

In our data, cooling regularized the spatial temperature patterns and reduced the discriminative contrast of most LZC variants and PC, consistent with the attenuation of most subject-level group differences under cooling, with thermogram-level descriptive summaries provided in Appendix Table E.11. This suggests that while structure-based approaches, such as hybrid segmentation of X-ray images [25], focus on permanent bone changes, thermal signals are highly sensitive to the immediate acquisition environment. In contrast, BPE retained sensitivity under cooling, indicating that evidence-theoretic handling of ties and near-equal temperatures can preserve informative ordinal structure when the thermal field becomes more homogeneous and sensor quantization effects become more pronounced.

The present results indicate that cooling does not primarily increase the global frequency of ties, but instead alters their spatial organization. Specifically, cooling induces locally homogeneous regions in which temperature gradients are reduced and ordinal relationships become less stable. In such regimes, deterministic permutation-based methods may introduce spurious variability due to unstable ranking, whereas BPE remains robust by explicitly modeling ordinal uncertainty. This distinction between tie frequency and spatial organization provides a more precise explanation for the observed protocol-dependent behavior of permutation-based descriptors.

Notably, under no-cooling conditions the thermal field exhibits higher baseline contrast, and BPE's conservative uncertainty penalty adds limited discriminative information under baseline conditions (Appendix Table E.10). Under cooling, BPE emerges as the dominant discriminator despite the absence of a global increase in near-tie prevalence (Table 4). This suggests that cooling may alter the spatial organization and local informativeness of near-equal values rather than simply increasing their overall frequency. This interpretation is consistent with the primary subject-level results and the thermogram-level descriptive trend shown in Appendix Table E.11.

4.3. Belief permutation entropy: Handling uncertainty and noise

Our results provide critical insights into the role of BPE in addressing sensor-related limitations. Unlike traditional permutation entropy, which forces a strict ranking even in the presence of ties, BPE utilizes Deng's entropy to explicitly penalize uncertainty. By distributing probability mass across focal sets, the algorithm accounts for the "ties" and quantization noise inherent in infrared sensors with limited thermal resolution. This transition from a purely probabilistic to an evidence-theoretic framework [19] allows the complexity analysis to remain informative even when spatial regularization under cooling increases the prevalence of near-equal temperatures. In our data, this tie-aware treatment proved critical: while BPE was less informative under baseline conditions due to a lower incidence of ties, it emerged as the dominant discriminator under cooling stress ($d \approx 0.80$). In this scenario, protocol-induced regularization increases the prevalence of near-equal temperatures, making BPE's evidence-theoretic handling of ties essential for preserving discriminative ordinal structures. This behavior should not be interpreted as a loss of sensitivity, but as methodological robustness: by penalizing tie-induced ambiguity, BPE avoids inflating apparent ordinal diversity due to noise, capturing instead the stable underlying spatial heterogeneity [19,31].

4.4. Clinical implications and comparison with texture analysis

Our findings align with previous research highlighting the value of entropy-based descriptors in rheumatology [24]. However, our results indicate that algorithmic and ordinal complexity measures provide complementary discrimination to classical texture descriptors (e.g., GLCM features reported in related thermography studies), particularly when symbolic encoding preserves the amplitude structure [25]. The proposed framework is best viewed as a methodological evaluation of encoding-dependent robustness rather than a ready-to-deploy diagnostic classifier. Under baseline conditions, the strongest single-feature effects remain modest (AUC $\approx$ 0.57–0.58), which limits stand-alone clinical utility. However, under cooling, BPE provides a robust univariate marker (AUC ≈ 0.71). This performance is particularly encouraging given that infrared thermography has recently been shown to correlate with subclinical Power Doppler synovitis even in joints that are not yet symptomatic [33]. Future work should integrate these descriptors into multivariate models to assess incremental value beyond established thermographic indices. Our finding that BPE under cooling stress provides a robust univariate marker aligns with the broader trend of developing energy-efficient digital biomarkers for decentralized rheumatological screening and point-of-care monitoring [2,34].

From a translational perspective, the proposed descriptors are computationally lightweight and can be extracted on standard CPU hardware without specialized acceleration, supporting their potential use in near-real-time thermographic screening workflows. However, they should currently be interpreted as complementary image-derived biomarkers quantifying spatial thermal heterogeneity, rather than substitutes for established rheumatological assessment tools such as DAS28, serological markers, or ultrasound. Their most plausible short-term role is as quantitative features within multimodal decision-support pipelines or longitudinal monitoring settings. Because these measures are computed directly from segmented thermograms within seconds per image and do not require model training, they are highly compatible with lightweight post-acquisition workflows. Unlike established intensity-based thermographic indices that quantify overall inflammatory burden through temperature distributions, the proposed complexity descriptors offer a distinct perspective by characterizing the spatial organization of the thermal field.

In practical terms, these descriptors are most naturally positioned as complementary features within thermographic workflows, for example in screening support, longitudinal monitoring of thermal heterogeneity, or protocol-sensitivity assessment rather than as stand-alone diagnostic tools.

Although the present study focuses on classification-oriented separability, the proposed complexity descriptors are not limited to binary discrimination. They may also be applicable to unsupervised settings such as clustering, anomaly detection, or longitudinal monitoring of thermal heterogeneity, which could provide additional insight into disease progression and treatment response.

5. Limitations

The study should be viewed as a methodological feasibility study rather than a definitive clinical validation. First, the cohort size was moderate and derived from a single-center acquisition setting, which may limit generalizability to broader and more heterogeneous populations, including different disease stages, treatment profiles, comorbidities, and demographic backgrounds. Although some effect sizes under cooling were substantial, external validation on larger, multi-center cohorts is required before any translational claims can be made. Moreover, the case-control design may overestimate apparent separability compared to real-world diagnostic settings involving heterogeneous differential diagnoses.

Second, the analysis relied on predefined parameter settings, including the multilevel quantization scheme and the embedding/tolerance parameters used for permutation-based descriptors. These settings were selected to ensure methodological consistency and avoid cohort-specific optimization, but alternative parameterizations may yield different absolute values. Although the supplementary sensitivity analysis showed that the qualitative ranking of descriptors remained stable, a more exhaustive hyperparameter study remains an important direction for future work.

Third, we focused specifically on baseline and cooling conditions and did not model potentially relevant clinical or physiological modifiers such as medication status, disease duration, vascular comorbidities, or peripheral circulation.

A further methodological limitation concerns the raster vectorization of thermograms. Although the validation analysis in Section 2.3 showed that row-transition artifacts are negligible and the traversal-robustness analysis preserved the qualitative ranking of descriptors, 1D vectorization still simplifies the native 2D topology of the thermal field. Future work should therefore explore 2D symbolic dynamics, graph-based formulations, or patch-wise complexity measures that operate directly on spatial domains.

Finally, although predictive validation was included through ROC analysis and univariate logistic regression, these results reflect only single-feature discrimination and should not be interpreted as a clinically deployable classifier. A larger prospective multi-center study will be required to establish clinically meaningful thresholds and to determine whether the proposed descriptors provide incremental value beyond established thermographic and rheumatological markers.

We did not assess probability calibration (e.g. Brier score or calibration curves), because the predictive component is intentionally limited to univariate discrimination (ROC–AUC) for methodological validation rather than clinical decision support.

Another limitation is that the present study was designed as a case-control discrimination analysis and did not evaluate correlations between complexity descriptors and established clinical disease-activity markers such as DAS28, CRP, RF, or ACPA. Consequently, the reported features should currently be interpreted as group-level discriminative descriptors rather than validated surrogates of clinical disease activity. Future studies should explicitly investigate whether these thermographic complexity measures track inflammatory burden, treatment response, or subclinical disease progression.

The response to the cooling protocol may also be influenced by subject-specific vascular and thermoregulatory factors beyond RA status itself, including baseline skin temperature, peripheral circulation, and vasoreactivity. Although the protocol was standardized and all recordings met predefined thermal-response criteria, such factors may contribute to residual variability in post-cooling thermal heterogeneity and should be explicitly modeled in future studies.

6. Conclusions

The present results show that the diagnostic value of information-theoretic complexity measures in hand thermography for rheumatoid arthritis is strongly shaped by both symbolic encoding design and the acquisition protocol. Under no-cooling conditions, several LZC variants and permutation complexity provided statistically significant but weak subject-level discrimination, indicating that some spatial heterogeneity differences are detectable at baseline but remain modest in magnitude. Under cooling, most LZC- and PC-based effects were attenuated, consistent with protocol-induced regularization of the thermal field.

In contrast, belief permutation entropy emerged as the strongest and most stable discriminator under cooling, supporting the use of tie-aware evidence-theoretic ordinal analysis when ties and near-ties become prevalent. These findings indicate that the main determinant of complexity-based thermographic performance is not the complexity metric alone, but the interaction between encoding choice, ordinal uncertainty handling, and acquisition protocol.

From a methodological perspective, the proposed features are computationally lightweight and reproducible, making them suitable as complementary biomarkers in future thermography-based decision-support pipelines. Future work should focus on multimodal integration, external multicenter validation, and correlation with established clinical activity markers to determine the translational value of these descriptors in rheumatoid arthritis screening and monitoring.

These findings should be interpreted as methodological evidence of how encoding strategy and ordinal uncertainty modeling influence complexity-based descriptors, rather than as a demonstration of clinically sufficient standalone diagnostic performance. The reported discrimination levels reflect subtle differences in spatial thermal heterogeneity and are intended primarily to support comparative evaluation of feature behavior under controlled acquisition protocols.

The additional near-tie analysis further indicates that this advantage is not explained by a global increase in near-tie frequency under cooling, but is more plausibly related to BPE's explicit representation of ordinal uncertainty in spatially regularized thermal patterns.

CRediT authorship contribution statement

Agnieszka Pregowska: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Jolanta Pauk:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Mikhail Ihnatouski:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Konrad Pauk:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis. **Janusz Szczepanski:** Writing – review & editing, Writing – original draft, Investigation, Funding acquisition, Conceptualization.

Funding

The study was partially supported by the Polish Ministry of Science and Higher Education as a part of the project WZ/WM–IIB/2/2024.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors express their thanks to Dr. Agnieszka Wasilewska for her contribution to thermography data acquisition.

Appendix A. Algorithmic details and pseudocode

This appendix provides an executable pseudocode for all steps used in the study. Each algorithm lists inputs, outputs, default parameters, and computational complexity.

Algorithm 1: A1. Thermography-to-Complexity Pipeline

Input: Thermogram $x \in \mathbb{R}^{H\times W}$, mask M, encoder $f(\cdot)$, parameters (m, τ, ϵ)
Output: Feature vector $\mathbf{z}$
Output definition: $\mathbf{z} = [\mathrm{LZC}, \mathrm{PC}_{\mathrm{norm}}, h_{\mathrm{BPE}}]$;
Preprocess: optional denoising/brightness normalization; hand segmentation to obtain M (per Sec. Image processing);
A1. $x \leftarrow x \odot M$ // apply mask **A2.** $s \leftarrow \mathrm{vec}(x \odot M)$ // row-major vectorization of ROI pixels, length L_{ROI} **A3.** $\tilde{s} \leftarrow f(s)$ // choose one of Algorithms A2 encoders **A4.** LZC $\leftarrow$ LZC_normalized($\tilde{s}$) // Algorithm A3 **A5.** $\mathrm{PC}_{\mathrm{norm}} \leftarrow \mathrm{PermutationComplexity}(s; m, \tau)$ // Algorithm A4 **A6.** $h_{\mathrm{BPE}} \leftarrow \mathrm{BeliefPermutationEntropy}(s; m, \tau, \epsilon)$ // Algorithm A5 **Return** $\mathbf{z}$;
Complexity: vectorization $O(HW)$; encoding $O(L)$; LZC $\tilde{O}(L)$; PC/BPE $O(L)$ for fixed (m, τ).

Algorithm 2: A2. Symbolic Encoders $f(\cdot)$

Input: Signal $s \in \mathbb{R}^{L_{\mathrm{ROI}}}$; thresholds Θ for multilevel
Output: $\tilde{s} \in \mathcal{A}^{L_{\mathrm{ROI}}}$
Binary: $\tilde{s}_t = \mathbb{I}[\, s_t \geq \mathrm{median}(s)\,]$, $\mathcal{A} = \{0, 1\}$;
Slope-direction: $\tilde{s}_t = \mathrm{sign}(s_t - s_{t-1}) \in \{-1, 0, 1\}$, with $\tilde{s}_1 = 0$;
Zero-crossing (mean-threshold): set baseline $\bar{s} = \frac{1}{L_{\mathrm{ROI}}} \sum_{t=1}^{L_{\mathrm{ROI}}} s_t$;
$\tilde{s}_t = \mathbb{I}[\, s_t \geq \bar{s}\,]$, $\mathcal{A} = \{0, 1\}$.
Multilevel: $\tilde{s}_t = k$ if $\Theta_{k-1} \leq s_t < \Theta_k$ for $k = 1, \ldots, K$ (e.g., Otsu or empirical quantiles);
Complexity: $O(L)$ (or $O(L \log K)$ if thresholds found via search).

Algorithm 3: A3. Normalized Lempel–Ziv Complexity (LZC)

Input: Symbolic sequence $\tilde{s} \in \mathcal{A}^L$, alphabet size $\alpha = |\mathcal{A}|$
Output: $c_\alpha(\tilde{s}) \in [0, 1]$
$C \leftarrow 1$; $i \leftarrow 1$; $k \leftarrow 1$;
while $i + k \leq L$ **do**
  if $\tilde{s}[i : i + k]$ *not seen in* $\tilde{s}[1 : i - 1]$ **then**
    $C \leftarrow C + 1$; $i \leftarrow i + k$; $k \leftarrow 1$;
  else
    $k \leftarrow k + 1$; **if** $i + k > L$ **then**
      $C \leftarrow C + 1$; break;
$c_\alpha(\tilde{s}) = \dfrac{C}{L/\log_\alpha L}$;
Complexity: $\tilde{O}(L)$ with hashing/suffix structures (near linear in practice).

Algorithm 4: A4. Permutation Complexity (PC)

Input: $s \in \mathbb{R}^L$; embedding dimension m; delay τ
Output: $\mathrm{PC}_{\mathrm{norm}} \in [0, 1]$
$N \leftarrow L - (m - 1)\tau$; initialize counts $h(\pi) = 0$ for all $\pi \in S_m$;
for $j = 1$ **to** N **do**
  $\mathbf{v} \leftarrow (s_j, s_{j+\tau}, \ldots, s_{j+(m-1)\tau})$;
  Obtain permutation π by ranking $\mathbf{v}$ with deterministic tie-break (e.g., by index);
  $h(\pi) \leftarrow h(\pi) + 1$;
$p(\pi) = h(\pi)/\sum_\pi h(\pi)$; $\mathrm{PC} = -\sum_\pi p(\pi) \log p(\pi)$; $\mathrm{PC}_{\mathrm{norm}} = \mathrm{PC}/\log(m!)$;
Complexity: $O(N\, m \log m)$; for small m this is linear in L.

Appendix B. Parameter specification and reproducibility tables

Algorithm 5: A5. Belief Permutation Entropy (BPE)

Input: $s \in \mathbb{R}^L$; (m, τ); tie tolerance ϵ
Output: $h_{\mathrm{BPE}} \in [0, 1]$
Initialize counts for focal sets $A \subseteq S_m$;
for $j = 1$ **to** $L - (m - 1)\tau$ **do**
  $\mathbf{v} \leftarrow (s_j, \ldots, s_{j+(m-1)\tau})$;
  Cluster components so that $|v_a - v_b| < \epsilon \Rightarrow$ same cluster;
  Derive the set A of permutations consistent with the partial order induced by clusters;
  $n(A) \leftarrow n(A) + 1$;
$m(A) = n(A)/\sum_A n(A)$;
$H_{\mathrm{BPE}} = -\sum_A m(A) \log\big(m(A)/(2^{|A|} - 1)\big)$; $h_{\mathrm{BPE}} = H_{\mathrm{BPE}}/\log(m!)$;
Complexity: $O(L)$ for fixed m, τ; enumeration of A is small for $m \leq 5$.

Algorithm 6: A6. Statistical Testing and Effect Size

Input: Features $\{\mathbf{z}_i\}$ from groups: Healthy (H), RA
Output: p-values (Welch t, Mann–Whitney U), Cohen's d
for *each measure* $\in \{$*LZC,* $\mathrm{PC}_{\mathrm{norm}}$, $h_{\mathrm{BPE}}\}$ *and each encoder* **do**
  Compute Welch two-sample t-test and Mann–Whitney U test;
  $d = \dfrac{\mu_{\mathrm{RA}} - \mu_{\mathrm{H}}}{\sqrt{\frac{\sigma_{\mathrm{RA}}^2 + \sigma_{\mathrm{H}}^2}{2}}}$;
  Store mean±SD, p-values, and d;
Complexity: $O(N)$ per measure/encoder.

Reproducibility checklist

To facilitate reproducibility, the following elements are fully specified:

- Data preprocessing and ROI extraction (Section 2.2)
- 2D-to-1D mapping strategy and validation (Section 2.3)
- Symbolic encoding schemes and parameters (Appendix Table B.6)
- Complexity computation algorithms (Algorithms A1–A5)
- Statistical testing, multiplicity correction, and predictive validation (Statistical Analysis and Predictive Validation subsection)
- Predictive validation protocol (Section 3.2)
- Computational cost and implementation details (Appendix D)
- Random seed control and aggregation strategy (subject-level median pooling)

All computations are deterministic given the specified parameters, except for randomized tie-breaking sensitivity analysis, where results are averaged over repeated realizations.

Appendix C. Rasterization validation/robustness

Resampling-based generalization assessment

To quantify the stability of thermogram-level (image-level) descriptive separability reported for comparability with prior work, we report 95% confidence intervals for Cohen's d using the large-sample variance (Hedges-Olkin) with image counts (Healthy n=291, RA n=186). The AUC-equivalent discrimination is obtained through normal-score mapping $\widehat{\mathrm{AUC}} = \Phi(d/\sqrt{2})$; CIs for $\widehat{\mathrm{AUC}}$ are calculated by transforming the d-CI endpoints. This provides a distributional, resampling-free assessment of stability at the effect-size level. As an optional sanity check, we outline repeated stratified subsampling (80/20 split, 100 repeats) on single-feature inputs; however, our primary claims rely on effect sizes and their CIs. Using a geometry-based diagnostic, the fraction of adjacency edges attributable to row-boundary jumps in

Table B.5
Parameters used in the analysis and default settings. Note: "multilevel" in Section 3 refers to the fourth encoding configuration evaluated in the traversal-robustness analysis and does not indicate a different alphabet size; all multilevel encodings used $K = 5$ quantization levels as specified above.

Category/Parameter	Value
Data and vectorization	
Image resolution	raw thermograms: 320 × 240 pixels (no spatial resizing applied)
Vector length L_{ROI}	number of ROI pixels after masking (variable across images)
Rasterization	row-wise (raster scan)
Encoders	
Binary threshold	global median of s
Slope-direction alphabet	$\{-1, 0, 1\}$ (diff with prepend)
Zero-cross baseline	sample mean $\bar{s}$; $\tilde{s}_i = \mathbb{I}[\, s_i \geq \bar{s}\,]$
Multilevel thresholds	$K = 5$; empirical quintiles (20/40/60/80%)
Complexity measures	
LZC normalization	$c = C/\left(L_{\mathrm{ROI}}/\log_\alpha L_{\mathrm{ROI}}\right)$, $\alpha = \lvert\mathcal{A}\rvert$
PC parameters	embedding $m = 5$, delay $\tau = 1$, normalization by $\log(m!)$
BPE parameters	$m = 5$, $\tau = 1$, tie tolerance $\epsilon = 10^{-3} \cdot \mathrm{IQR}(s)$
Statistical analysis	
Tests	Welch t, Mann–Whitney U
Effect size	Cohen's d

Table B.6
Complete list of parameters used for symbolic encoding and complexity computation (reproducibility specification).

Parameter	Value/Definition
Image normalization	$s^* = (s - \min s)/(\max s - \min s)$
Vector length L_{ROI}	number of ROI pixels after masking (variable across images); e.g., up to 320 × 240 for full frame
Binary encoding	
Threshold	global median of s^*
Alphabet	$\{0, 1\}$
Slope-direction encoding	
Rule	$\mathrm{sign}(s^*_{i+1} - s^*_i)$
Alphabet	$\{-1, 0, 1\}$
Zero-crossing encoding	
Baseline	sample mean $\bar{s}$
Rule	$\tilde{s}_i = \mathbb{I}[\, s_i \geq \bar{s}\,]$
Alphabet	$\{0, 1\}$ (below/above mean)
Multilevel encoding	
Number of levels K	5
Thresholds	empirical quantiles $t_k = Q(k/K)$, $k = 0, \ldots, K$
Mapping rule	$t_{k-1} \leq s^*_i < t_k \Rightarrow \tilde{s}_i = k, \quad k = 1, \ldots, K$
Alphabet size α	$K = 5$
Lempel–Ziv complexity	
Normalization	$c = C/(L_{\mathrm{ROI}}/\log_\alpha L_{\mathrm{ROI}})$
Permutation complexity (PC)	
Embedding dimension m	5
Delay τ	1
Normalization	$PC/\log(m!)$
Belief permutation entropy (BPE)	
Embedding dimension m	5
Tie tolerance ϵ	$10^{-3} \cdot \mathrm{IQR}(s^*)$
Entropy	Deng's entropy: $H_{\mathrm{BPE}} = -\sum_A m(A) \log_2\left(\frac{m(A)}{2^{\lvert A\rvert}-1}\right)$

raster scanning was 0.1562% (median; IQR 0.0000) in both protocols (Table C.7), i.e., well below 0.5%. This confirms that the raster boundary introduces a negligible proportion of artificial adjacencies compared to the total number of 4-neighborhood adjacencies inside the ROI. Subject-level confidence intervals are reported in Section 3.4 and Table 1. The traversal-robustness subset was intentionally restricted to a reduced feature set and should not be interpreted as a comparative ranking benchmark across all descriptors. Therefore, the induced 1D representation preserves the statistical structure relevant for complexity measures without introducing measurable bias. Note that the AUC values reported in this robustness analysis are not directly comparable to the primary protocol-matched results presented in Table E.12. The robustness analysis is based on a reduced feature set and serves only to assess the stability of descriptor ranking under alternative traversal strategies, rather than to provide primary estimates of classification performance.

Justification of raster vectorization

Although thermograms are inherently two-dimensional, all complexity measures examined in this study (LZC, PC, BPE) are defined for symbolic one-dimensional sequences. Consequently, a deterministic traversal of the image domain is required. We used standard row-wise raster scanning, which preserves all pixel intensities and maintains local adjacency along a continuous space-filling path. This approach is widely adopted in complexity-based image analysis, including LZC studies on radiographs and ultrasound B-mode textures. Because thermal images exhibit smooth spatial gradients and large size (320 × 240), the potential boundary effects at row transitions represent less than 0.4% of the total sequence length and therefore negligibly influence the resulting symbolic dynamics. For completeness, we also evaluated a patch-based alternative (non-overlapping 8 × 8 blocks encoded independently using the same symbolic scheme). The resulting effect sizes (Healthy *versus* RA) were consistent with the raster-based results

Table C.7
Quantitative diagnostics of robustness of 2D-to-1D mapping across traversal strategies. Values are reported as cohort medians with interquartile ranges (IQR) across images.

Diagnostic	No-cooling	Cooling
Row-boundary adjacency share (2D) [%]	0.1562 (IQR: 0.0000)	0.1562 (IQR: 0.0000)
Spearman ρ (row vs column) for PC and BPE (cohort-level)	0.251	0.197
Jensen–Shannon divergence of ordinal-pattern histograms (row vs column)	0.01812 (IQR: 0.01216)	0.02001 (IQR: 0.01320)

Table D.8
Runtime and memory per image (prototype single-threaded Python implementation).

Encoder	Time [ms/img]	Peak mem [MB]	Asymptotic (time/mem)
Binary	9664.35	63.29	$O(L)/O(1)$
Slope	4951.43	0.35	$O(L)/O(1)$
Zero-crossing	10 415.23	0.35	$O(L)/O(1)$
Multilevel (K=5)	7369.18	0.36	$O(L+K)/O(1)$

Notes. (i) Times reflect a pure-Python, single-pass reference; vectorized/C-accelerated implementations are expected to be an order of magnitude faster while preserving linear time in the raster length $L = H \times W$. (ii) The elevated peak for *Binary* likely stems from prototype-level allocation/measurement (e.g., missing per-image reset of `tracemalloc`); other encoders show the expected sub-MB footprint dominated by the image buffer.

Table E.9
Ablation summary using effect sizes (d) and AUC proxy $\widehat{\mathrm{AUC}} = \Phi(|d|/\sqrt{2})$.

Protocol	Feature	$\lvert d\rvert$	$\widehat{\mathrm{AUC}}$
No-cooling	PC	0.289	0.581
No-cooling	LZC_zero_cross	0.269	0.576
No-cooling	LZC_slope	0.267	0.575
No-cooling	LZC_binary	0.233	0.565
No-cooling	LZC_multilevel	0.191	0.554
No-cooling	BPE	0.011	0.503
Cooling	BPE	0.795	0.713
Cooling	LZC_zero_cross	0.340	0.595
Cooling	LZC_multilevel	0.251	0.571
Cooling	LZC_slope	0.165	0.546
Cooling	LZC_binary	0.140	0.539
Cooling	PC	0.089	0.525

and preserved the same qualitative conclusions regarding protocol dependence and encoding sensitivity. These findings support the notion that raster vectorization does not materially bias the conclusions.

Appendix D. Efficiency and computational cost

Efficiency analysis: time, memory and algorithmic complexity

We report wall-clock time per image and peak memory for each encoder, together with asymptotic costs. All analyses were performed in Python 3.12.4 using the SciPy 1.13.1 and NumPy 1.26.4 libraries. Values are medians across all images (single-threaded reference implementation). The reported timings include diagnostic overhead (e.g., memory tracing) and are not indicative of optimized implementations.

The reported runtimes correspond to a prototype single-threaded reference implementation intended to document algorithmic order-of-growth rather than optimized performance. Absolute wall-clock times may be inflated by Python-level overhead or profiling instrumentation (e.g., memory tracing). Accordingly, the key reproducibility result is the linear dependence on the ROI raster length L_{ROI} for fixed embedding parameters (m, τ), rather than the absolute timing values.

Multilevel encoding introduces only a minor overhead relative to binary and slope encodings, with all methods remaining effectively linear-time in image size. Memory usage is dominated by the raster buffer and remains below 1 MB for all encoders.

Appendix E. Additional results

Ablation study: Effect-size and AUC-proxy benchmarking

To quantify the relative contribution of protocol choice and feature definition to group discrimination, we performed a lightweight ablation analysis based on standardized effect sizes. For each feature, we report Cohen's d (Tables E.9–E.11) and its corresponding AUC proxy using the Gaussian-score mapping

$$\widehat{\mathrm{AUC}} = \Phi\left(\frac{|d|}{\sqrt{2}}\right), \tag{E.1}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. This mapping provides an interpretable link between separability and expected ROC performance in the univariate setting. The ablation results show that under no-cooling conditions, the strongest single-feature discriminators remain weak (typical $\widehat{\mathrm{AUC}} \approx 0.56$–$0.58$), with PC and LZC computed from slope/zero-crossing encodings yielding the highest separability. Under cooling stress, most LZC-based effects are attenuated, while BPE becomes dominant, reaching $\widehat{\mathrm{AUC}} \approx 0.71$, consistent with the ROC–AUC estimates in Table E.12.

Parameter sensitivity analysis

To assess robustness to parameterization, we performed additional analyses under small perturbations of key parameters. For BPE, embedding dimensions $m \in \{4, 5\}$ and tolerance values $\epsilon \in \{5 \times 10^{-4}, 10^{-3}, 2 \times 10^{-3}\} \cdot \mathrm{IQR}(s^*)$ were evaluated. For multilevel encoding, the number of quantization levels was varied ($K \in \{4, 5, 6\}$). Across these settings, absolute feature values varied modestly, but the qualitative ranking of descriptors remained unchanged. Under cooling conditions, BPE consistently provided the strongest discrimination, confirming that the main conclusions are not driven by specific parameter choices. In particular, no alternative setting reversed the main protocol-dependent conclusion that BPE is the strongest descriptor under cooling.

Hardware/software requirements and cost

Acquisition hardware. Infrared camera as described in Methods (FLIR E60bx, 320 × 240). No special compute hardware during acquisition.

Compute. CPU-only; no GPU required. All feature computations are linear in the ROI raster length L_{ROI} for fixed (m, τ); in a prototype single-threaded Python implementation this corresponds to seconds per image on a standard laptop/desktop (e.g., 4-8 cores, ≥8 GB RAM), while vectorized or compiled implementations are expected to reduce runtime substantially without changing the asymptotic cost.

Software. Open-source Python stack (NumPy, SciPy, matplotlib). No paid licenses are required to reproduce the analysis.

Data availability

Data will be made available on request.

Table E.10
Appendix thermogram-level descriptive comparison of complexity measures between Healthy and RA groups under **No-cooling (NC)**. These summaries are retained for comparability with earlier thermography studies only; primary inferential conclusions are based on subject-level analyses. Values are mean ± SD. *p*-values are from the Mann–Whitney *U* test and BH-FDR adjusted within the NC feature family. Significant values ($p_{FDR} < 0.05$) are highlighted in bold.

Measure	Encoding/Type	Healthy	RA	p_{FDR}	Cohen's *d*
LZC	Binary	0.306 ± 0.098	0.282 ± 0.104	**0.0076**	−0.233
	Slope	0.847 ± 0.037	0.837 ± 0.033	**0.0076**	−0.267
	Zero-crossing	0.281 ± 0.073	0.262 ± 0.064	**0.0095**	−0.269
	Multilevel	0.380 ± 0.057	0.370 ± 0.051	0.0602	−0.191
PC	Continuous (no encoding)	0.681 ± 0.023	0.687 ± 0.022	**0.0095**	0.289
BPE	Continuous (no encoding)	0.187 ± 0.043	0.187 ± 0.037	0.8405	0.011

LZC is computed from symbolized sequences and therefore depends on the encoding strategy. In contrast, permutation complexity (PC) and belief permutation entropy (BPE) are computed directly from the continuous normalized raster signal (embedding m=5, τ=1), and therefore appear as single predictors in the comparison table. Accordingly, BH-FDR correction applies only to the four LZC encoding variants within each protocol, while PC and BPE are treated as standalone continuous-signal descriptors rather than encoding families.

Table E.11
Appendix thermogram-level descriptive comparison of complexity measures between Healthy and RA groups under **Cooling**. These summaries are retained for comparability with earlier thermography studies only; primary inferential conclusions are based on subject-level analyses. Values are mean ± SD. *p*-values are from the Mann–Whitney *U* test and BH-FDR adjusted within the NC feature family. Significant values ($p_{FDR} < 0.05$) are highlighted in bold.

Measure	Encoding/Type	Healthy	RA	p_{FDR}	Cohen's *d*
LZC	Binary	0.210 ± 0.060	0.219 ± 0.069	0.4367	0.140
	Slope	0.786 ± 0.039	0.780 ± 0.035	0.1971	−0.165
	Zero-crossing	0.211 ± 0.038	0.200 ± 0.031	0.0641	−0.340
	Multilevel	0.312 ± 0.036	0.321 ± 0.038	0.1971	0.251
PC	Continuous (no encoding)	0.637 ± 0.024	0.635 ± 0.020	0.3095	−0.089
BPE	Continuous (no encoding)	0.163 ± 0.033	0.191 ± 0.037	**0.000002**	0.795

LZC is computed from symbolized sequences and therefore depends on the encoding strategy. PC and BPE are computed directly from the continuous normalized raster signal (embedding m=5, τ=1), and therefore appear as single predictors. BH-FDR correction applies to the four LZC encoding variants only. PC and BPE are continuous-signal descriptors and are therefore reported as single predictors rather than encoding families.

Table E.12
Predictive validation stability for univariate logistic regression models (Healthy *versus* RA), evaluated at the **subject level** after median aggregation of thermogram-level features within each participant. AUC is reported as the bootstrap mean (1000 subject-level resamples) with 95% CI and as grouped subject-wise cross-validation (GroupCV; repeated 5-fold CV).

Protocol	Feature	Cohort	AUC (bootstrap mean)	95% CI	AUC (GroupCV mean ± SD)
No-cooling	LZC_binary	Full subject cohort	0.582	[0.527, 0.639]	0.581 ± 0.060
No-cooling	LZC_slope	Full subject cohort	0.581	[0.529, 0.631]	0.580 ± 0.022
No-cooling	PC	Full subject cohort	0.575	[0.527, 0.628]	0.574 ± 0.041
No-cooling	LZC_zero_cross	Full subject cohort	0.573	[0.522, 0.623]	0.576 ± 0.027
No-cooling	LZC_multilevel	Full subject cohort	0.552	[0.500, 0.603]	0.556 ± 0.047
No-cooling	BPE	Full subject cohort	0.496	[0.439, 0.546]	0.463 ± 0.020
Cooling	BPE	Matched subject cohort	0.709	[0.639, 0.778]	0.713 ± 0.072
Cooling	LZC_zero_cross	Matched subject cohort	0.595	[0.512, 0.668]	0.592 ± 0.078
Cooling	LZC_multilevel	Matched subject cohort	0.568	[0.487, 0.651]	0.574 ± 0.128
Cooling	LZC_slope	Matched subject cohort	0.560	[0.483, 0.640]	0.565 ± 0.054
Cooling	PC	Matched subject cohort	0.545	[0.464, 0.622]	0.468 ± 0.112
Cooling	LZC_binary	Matched subject cohort	0.533	[0.454, 0.619]	0.533 ± 0.084

References

[1] Y. Cai, X. Zhang, X. Wang, F. Yu, K. Tang, X. Xu, et al., The burden of rheumatoid arthritis: Findings from the 2019 global burden of diseases study and forecasts for 2030 by Bayesian age-period-cohort analysis, J. Clin. Med. 12 (4) (2023) 1291, http://dx.doi.org/10.3390/jcm12041291.

[2] Y. Ma, X. Li, X. Xu, X. Zhang, F. Yu, K. Tang, et al., Global, regional and national burden of rheumatoid arthritis (RA): current estimates and future projections based on GBD 2021, Biomark Res. 13 (2025) 47, http://dx.doi.org/10.1186/s40364-025-00760-8.

[3] Z. Zhang, X. Gao, S. Liu, Q. Wang, Y. Wang, S. Hou, J. Wang, Y. Zhang, Global, regional, and national epidemiology of rheumatoid arthritis among people aged 20–54 years from 1990 to 2025, Sci. Rep. 15 (2025) 10736, http://dx.doi.org/10.1038/s41598-025–92150–1.

[4] F. Yu, H. Chen, Q. Li, M. Tao, Z. Jin, L. Geng, L. Sun, Secular trend of mortality and incidence of rheumatoid arthritis in global 1990–2019: an age period cohort analysis and joinpoint analysis, BMC Pulm. Med. 23 (1) (2023) 356, http://dx.doi.org/10.1186/s12890-023-02594-2.

[5] R.J. Stack, P. Nightingale, C. Jinks, K. Shaw, Herron-Marx S, R. Horne, C. Deighton, P. Kiely, C. Mallen, K. Raza, DELAY study syndicate. Delays between the onset of symptoms and first rheumatology consultation in patients with rheumatoid arthritis in the UK: an observational study, BMJ Open. 9 (3) (2019) e024361, http://dx.doi.org/10.1136/bmjopen-2018-024361.

[6] Y. Jiang, S. Zhong, S. He, J. Weng, L. Liu, Y. Ye, H. Chen, Biomarkers (mrnas and non-coding RNAs) for the diagnosis and prognosis of rheumatoid arthritis, Front. Immunol. 14 (2023) 1129383, http://dx.doi.org/10.3389/fimmu.2023.1129383.

[7] Renaudineau Y, Advancements in early diagnosis of rheumatoid arthritis: a shift toward precision medicine, Int. J. Clin. Rheumatol. 30, 19 (10) (2024) 267–270, http://dx.doi.org/10.37532/1758-4272.2024.19(10).267-270.

[8] Ruiz-Romero C, Fernández-Puente P, González L, A. Illiano, L. Lourido, R. Paz, P. Quaranta, Perez-Pampín E, González A, F.J. Blanco, V. Calamia, Association of the serological status of rheumatoid arthritis patients with two circulating protein biomarkers: a useful tool for precision medicine strategies, Front. Med (Lausanne). 28 (9) (2022) 963540, http://dx.doi.org/10.3389/fmed.2022.963540.

[9] B. Nakken, G. Papp, V. Bosnes, M. Zeher, G. Nagy, P. Szodoray, Biomarkers for rheumatoid arthritis: from molecular processes to diagnostic applications-current concepts and future perspectives, Immunol. Lett. 189 (2017) 13–18, http://dx.doi.org/10.1016/j.imlet.2017.05.010.

[10] I. Minopoulou, A. Kleyer, Yalcin-Mutlu M, F. Fagni, S. Kemenes, C. Schmidkonz, et al., Imaging in inflammatory arthritis: progress towards precision medicine,

Nat. Rev. Rheumatol. 19 (10) (2023) 650–665, http://dx.doi.org/10.1038/s41584-023-01016-1.
[11] A. Gupta, S. Anis, P. de Pablo, Imaging tests as predictors of progression to rheumatoid arthritis in clinically suspect arthralgia:a systematic review and metaanalysis, Rheumatol. (Oxford) 64 (6) (2025) 3255–3265, http://dx.doi.org/10.1093/rheumatology/keaf045.
[12] M.H. Abdelbary, A.S. Khidr, A.O. Kamel, Hand and wrist magnetic resonance imaging versus high resolution ultrasonography in patients with rheumatoid arthritis, Egypt J. Radiol. Nucl. Med. 56 (2025) 107, http://dx.doi.org/10.1186/s43055-025-01529-7.
[13] J. Pauk, A. Wasilewska, M. Ihnatouski, Infrared thermography sensor for disease activity detection in rheumatoid arthritis patients, Sensors 19 (2019) 3444, http://dx.doi.org/10.3390/s19163444.
[14] I. Morales-Ivorra, O. Taverner, S. Castell, D. Fischer, P. Martínez-Osuna, C. Battioui, M.A. Marín-López, External validation of the machine learning-based thermographic indices for rheumatoid arthritis: a prospective longitudinal study, Diagnostics (Basel) 30, 14 (13) (2024) 1394, http://dx.doi.org/10.3390/diagnostics14131394.
[15] H. Zhao, J. Xie, Y. Chen, J. Cao, W.H. Liao, H. Cao, Diagnosis of neurodegenerative diseases with a refined lempel–ziv complexity, Cogn. Neurodyn. 18 (3) (2024) 1153–1166, http://dx.doi.org/10.1007/s11571-023-09973-9.
[16] A. Lempel, J. Ziv, On the complexity of finite sequences, IEEE Trans. Inf. Theory. 22 (1) (1976) 75–81, http://dx.doi.org/10.1109/TIT.1976.1055501.
[17] A. Pregowska, A. Casti, E. Kaplan, E. Wajnryb, J. Szczepanski, Information processing in the LGN: a comparison of neural codes and cell types, Biol. Cybernet. 113 (4) (2019) 453–464, http://dx.doi.org/10.1007/s00422-019-00801-0.
[18] A. Pregowska, K. Proniewska, P. van Dam, J. Szczepanski, Using Lempel–Ziv complexity as effective classification tool of the sleep-related breathing disorders, Comput. Methods Programs Biomed. 182 (2019) 105052–1–7, http://dx.doi.org/10.1016/j.cmpb.2019.105052.
[19] J. Xie, G. Xu, X. Chen, X. Zhang, R. Chen, Z. Yang, B. Li, S. Zhang, Belief permutation entropy of time series: a natural transition in analytical framework from probability theory to evidence theory, Inf. Sci. (N Y) 718 (2025) 122352, http://dx.doi.org/10.1016/j.ins.2025.122352.
[20] Y.J. Huang, C.H. Lin, S. Miao, K. Zheng, L. Lu, Y. Lu, C. Lin, C.F. Kuo, Radiographic bone texture analysis using deep learning models for early rheumatoid arthritis diagnosis, J. Imaging Inf. Med. (2025) http://dx.doi.org/10.1007/s10278-025-01579-3.
[21] R.K. Ahalya, F.M. Almutairi, U. Snekhalatha, et al., RANet: a custom CNN model and quanvolutional neural network for the automated detection of rheumatoid arthritis in hand thermal images, Sci. Rep. 13 (2023) 15638, http://dx.doi.org/10.1038/s41598-023-42111-3.
[22] I. Morales-Ivorra, J. Narváez, C. Gómez-Vaquero, C. Moragues, J.M. Nolla, J.A. Narváez, M.A. Marín-López, A thermographic disease activity index for remote assessment of rheumatoid arthritis, RMD Open. 8 (2) (2022) e002615, http://dx.doi.org/10.1136/rmdopen-2022-002615.
[23] S. Uğur, Y. İrim, Ayça Yücel A., H.F. Carlak, C. Ka¸car, A comparison of thermal characteristics of the small joints of the hands between patients with rheumatoid arthritis and healthy controls, Arch Rheumatol 12, 39 (4) (2024) 617–623, http://dx.doi.org/10.46497/ArchRheumatol.2024.10753.
[24] U. Snekhalatha, T. Rajalakshmi, M. Gopikrishnan, N. Gupta, Computer-based automated analysis of X-ray and thermal imaging of knee region in evaluation of rheumatoid arthritis, Proc. Inst. Mech. Eng. H 231 (12) (2017) 1178–1187, http://dx.doi.org/10.1177/0954411917737329.
[25] G. Rajesh, N. Malarvizhi, M.F. Leung, A hybrid segmentation algorithm for rheumatoid arthritis diagnosis using X-ray images, Big Data Cogn. Comput. 8 (9) (2024) 104, http://dx.doi.org/10.3390/bdcc8090104.
[26] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88 (17) (2002) 174102, http://dx.doi.org/10.1103/PhysRevLett.88.174102.
[27] J.M. Amigó, M.B. Kennel, L. Kocarev, The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems, Phys. D. 210 (1-2) (2005) 77–95, http://dx.doi.org/10.1016/j.physd.2005.07.006.
[28] T. Bossomaier, L. Barnett, M. Harr, J.T. Lizier, An Introduction To Transfer Entropy: Information Flow in Complex Systems, first ed., Springer, New York, 2016, http://dx.doi.org/10.1007/978-3-319-43222-9.
[29] J. Hu, J. Gao, J.C. Principe, Analysis of biomedical signals by the Lempel–Ziv complexity: the effect of finite data size, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2606–2609, http://dx.doi.org/10.1109/TBME.2006.883825.
[30] R. Nagarajan, J. Szczepanski, E. Wajnryb, Interpreting non-random signatures in biomedical signals with Lempel–Ziv complexity, Phys. D: Nonlinear Phenom. 237 (3) (2008) 359–364, http://dx.doi.org/10.1016/j.physd.2007.09.007.
[31] H. Azami, J. Escudero, Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings, Biomed. Signal Process. Control. 23 (1) (2016) 28–41, http://dx.doi.org/10.1016/j.bspc.2015.08.004.
[32] Y. Deng, Deng entropy, Chaos Solitons Fractals 91 (2016) 549–553, http://dx.doi.org/10.1016/j.chaos.2016.07.014.
[33] Y.K. Tan, J. Thumboo, Understanding ultrasound power Doppler synovitis at clinically quiescent joints and thermographic joint inflammation assessment in patients with rheumatoid arthritis, Diagn. (Basel). 14 (21) (2024) 2384, http://dx.doi.org/10.3390/diagnostics14212384.
[34] S. Bhowmick, A. Saha, S. Deb, A. De, A. Srivastava, Medical image classification using lightweight deep spiking neural network, Iran J. Sci. Technol. Trans. Electr. Eng. 49 (2025) 589–600, http://dx.doi.org/10.1007/s40998-025-00808-3.